Regulatory Sandbox Final Report: FutureFlow

A summary of FutureFlow's participation in the ICO's Regulatory Sandbox Beta

Date: October 2020

Information Commissioner's Office

Contents

1. Introduction 3
2. Executive summary 5
3. Platform description 7
4. Key data protection considerations 16
5. Ending statement 25

1. Introduction

1.1. FutureFlow Research Inc ("**FutureFlow"**) provides a Transaction Monitoring and Forensic Analytics Platform which monitors the flow of funds in the financial system. The platform enables financial institutions to contribute transactional data which has undergone pseudonymisation in bulk, enabling multiple financial institutions, Regulators, and agencies to work together to detect and ultimately tackle electronic financial crime.

1.2. This collaborative approach to fighting financial crime opens up the prospect of higher detection rates with lower false positives, while reducing the burden of scrutiny on each individual and business consumer.

1.3. As part of FutureFlow's engagement with the Sandbox, we have set the following four objectives:

- **Objective 1:** FutureFlow and the ICO will work to map the ways in which FutureFlow's data could flow given their proposed operating models, and clarify which parties are undertaking key data protection roles in respect of the data utilised by FutureFlow's platform (ie who is a Controller, Joint Controller and/or Processor).
- **Objective 2:** FutureFlow and the ICO will develop a system to effectively quantify and manage the risk of reidentification of personal data in FutureFlow's platform.
- **Objective 3:** FutureFlow and the ICO will develop relevant collateral documentation to help ensure that FutureFlow process personal data in compliance with UK Data Protection legislation.
- **Objective 4:** FutureFlow, with guidance from the ICO should implement sufficient measures to ensure the ongoing security and integrity of the data they process.

1.4. FutureFlow were accepted into the Sandbox on 01 July 2019 and a Senior Case Officer was appointed. The ICO attended the offices of FutureFlow on 30 July 2019 to conduct a scoping visit to aid the drafting of FutureFlow's Sandbox Plan.

1.5. The content of the Sandbox Plan was agreed by Vadim Sobolevski of FutureFlow on 20 August 2019 and was subsequently approved by the ICO Sandbox commissioning and advisory group on 09 September 2019.

1.6. On 21 August 2020 FutureFlow and the ICO completed the last piece of work detailed in FutureFlow's Sandbox plan, bringing FutureFlow's participation in the ICO's Regulatory Sandbox beta to an end.

2. Executive summary

2.1. FutureFlow's Sandbox Plan focused around their platform's use of data analytics to detect, identify and ultimately tackle instances of financial crime, including money laundering. The platform itself leverages transactional data provided by multiple financial institutions, such as banks, to create a map of the financial network, which is then scrutinised to identify behaviours which could be potentially viewed as suspicious. FutureFlow have developed two modes of operating while developing their product, the first (Direct Mode) involves FutureFlow's clients sending personal data which has already undergone pseudonymisation directly to FutureFlow for processing. The second operating model (Indirect Mode) involves FutureFlow's clients sending data which has already undergone pseudonymisation to a Trusted Third Party for de-duplication[1], cleaning, further pseudonymisation etc.[2] before FutureFlow processes the data.

2.2. During FutureFlow's participation in the Sandbox we have considered the following key data protection issues:

[1] See 3.7 for an explanation of this process.
[2] Further information about FutureFlow's Direct and Indirect modes of operation is available in the 'Product Description' section of this report.

- **Complex data controllership issues** (ie which party should be considered a Controller, Joint Controller or Processors with regard to the processing activity in Direct Mode and how would this change in Indirect Mode?) – the ICO reached the view that in the context of FutureFlow's processing activity in both the Direct Mode and Indirect Mode of operation, FutureFlow would likely be considered a Processor, acting on the behalf of their clients, who would likely be considered the Controllers. When operating in the Indirect Mode, the ICO reached the view that the Trusted Third Party should likely be considered a Processor[3] also.

- **Could the data processed by FutureFlow be considered anonymous?** – The ICO reached the view that the data processed via FutureFlow's platform would likely not be considered anonymous, despite the effectiveness of the pseudonymisation techniques in reducing the identifiability of the data, this is because, despite being heavily pseudonymised (particularly when utilising the services of a Trusted Third Party in FutureFlow's Indirect Mode of operation), the risk of re-identification by a motivated intruder would still be regarded as reasonably likely if such an intruder gained access to the pseudonymised data.

[3] Please Note: When operating in Indirect Mode FutureFlow would not be considered a sub processor for the Trusted Third Party, this is because FutureFlow and the Trusted Third Party perform distinct processing activities independent of one another. Further detail on this point is provided in the 'Key Data Protection Considerations' section of this document.

- **How should FutureFlow look to comply in a general sense with the UK Data Protection Legislation?** – The ICO worked with the organisation during their time in the Sandbox to build their data protection capacity by offering ad hoc advice to the organisation and assisting them in drafting suitable collateral documentation to support their ongoing compliance – this documentation included a Record of Processing Activity (ROPA), a Data Protection Policy and a model Data Protection Impact Assessment (DPIA).

3. Platform description

3.1. The platform provided by FutureFlow utilises data on financial transactions that contain personal data which has undergone pseudonymisation. This transactional information is received from various institutions, including:

- Banks and other financial institutions; and
- Trusted third parties who have collated, deduplicated and otherwise 'cleaned' personal data which relates to financial transactions. This data will have been received by the Trusted Third Parties from FutureFlow's clients, who are predominantly banks and other financial institutions.

3.2. The financial transaction data which FutureFlow will process includes the following categories of personal data[4]:

- Account identifier (such as account number, sort code, IBAN number, etc);
- Transaction value(s);
- Transaction IDs; and
- Time-stamps.

[4] Please Note: FutureFlow's platform does not process special category personal data as defined by Article 9 of the GDPR.

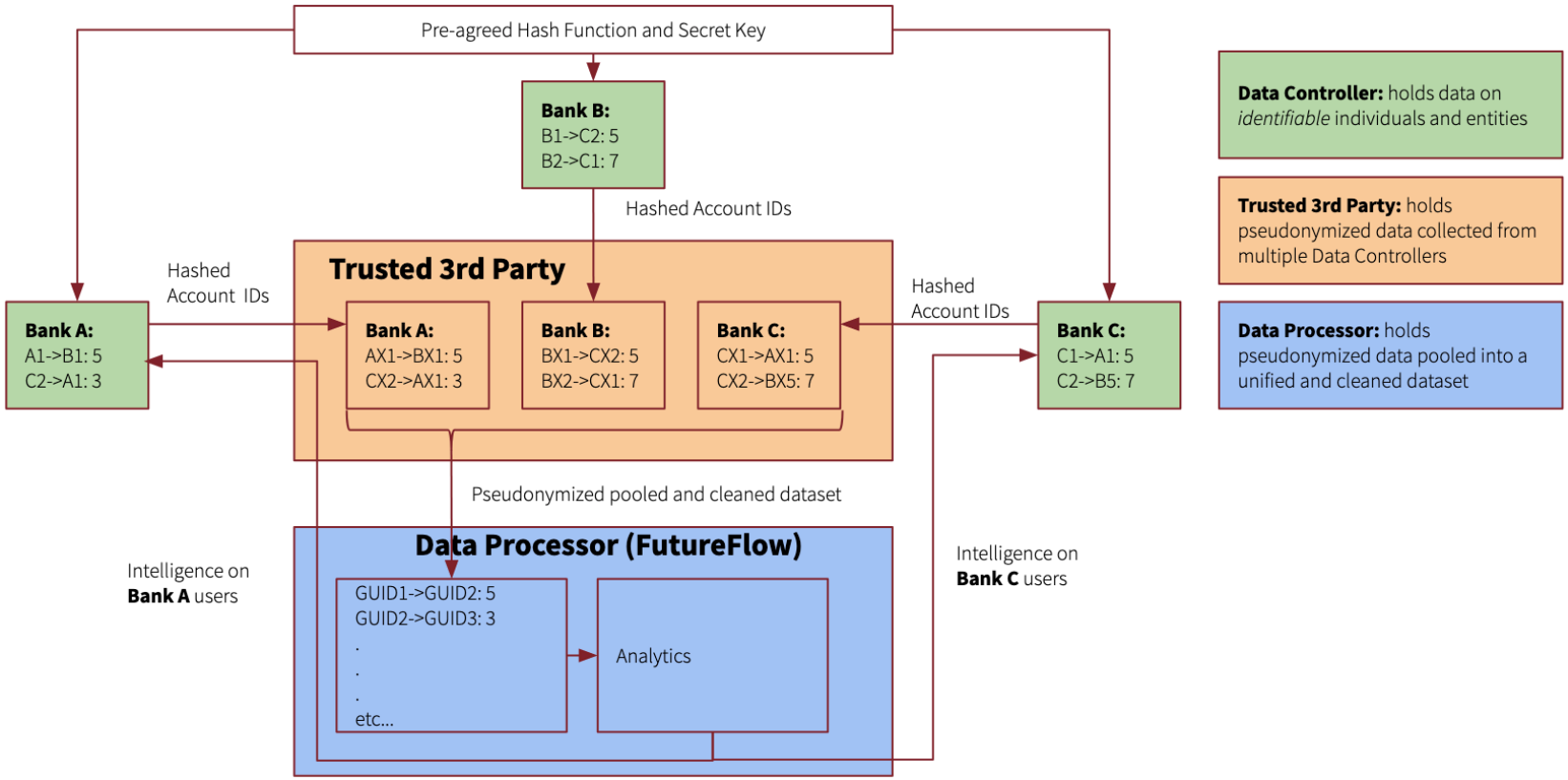
3.3. FutureFlow's platform can operate as a cross-bank Anti-Money-Laundering Utility in two modes:

- **Indirect Mode:** which involves FutureFlow receiving transactional data with account identifiers from a Trusted Third Party who first receives transactional data which has already undergone pseudonymisation from FutureFlow's clients such as banks and other financial institutions.
- **Direct Mode:** which involves FutureFlow receiving transactional data including account identifiers directly from their clients who should have already utilised pseudonymisation techniques on the data before sending it to FutureFlow.

3.4. While the core of FutureFlow's processing remains the same, the two modes differ in terms of how the data on which FutureFlow operates is first pre-processed. While the Indirect Mode introduces additional coordination challenges, it offers a more meaningful level of obfuscation (as described further below), thus reducing the level of risk to individuals.

Indirect Mode

3.5. In the Indirect Mode a Trusted Third Party facilitates the exchange of data between each financial institution client and FutureFlow. Suitable Trusted Third Parties are likely to be large consultancies, system integrators, or other organisations of a similar stature that command a trusting business relationship with multiple banks and financial institutions by the nature of their business. For example, in the first pilot conducted by FutureFlow using real-life data, a Big-4 consultancy firm served as a de-facto Trusted Third Party. An illustration of how FutureFlow anticipates this transaction will take place is provided below:



3.6. In the Indirect Mode, the Trusted Third Party helps multiple banks to coordinate and agree on a common convention for identifying and pseudonymising account identifiers in the transactional datasets, which will subsequently be processed by FutureFlow. As described in the above diagram, the process includes:

- **Step 1** - Agreeing a common convention on identifying the sending and receiving account in a transaction (such as using an IBAN, a combination of Account Number and Sort Code, etc);

- **Step 2** - Agreeing a common Hash Function for pseudonymisation. The chosen Hash Function should be of industry-recognized strength and should conform to the standard principles of hashing:

    A. **One-way:** a hash is irreversible to the original value computationally (except by trial and error)

    B. **Consistent:** hashing the same value always produces the same hash; and

    C. **Collision-Free:** hashing two different values to the same hash is highly unlikely, and

- **Step 3** - Agreeing a common Secret Key and a common convention of mixing this Secret Key with the account identifier agreed in Step 1 above[5]

3.7. Each participating bank submits its transactional dataset, with the account identifiers already having undergone pseudonymisation according to the steps outlined above, to the Trusted Third Party. The Trusted Third Party then combines multiple datasets into a pooled dataset and performs the following transformations:

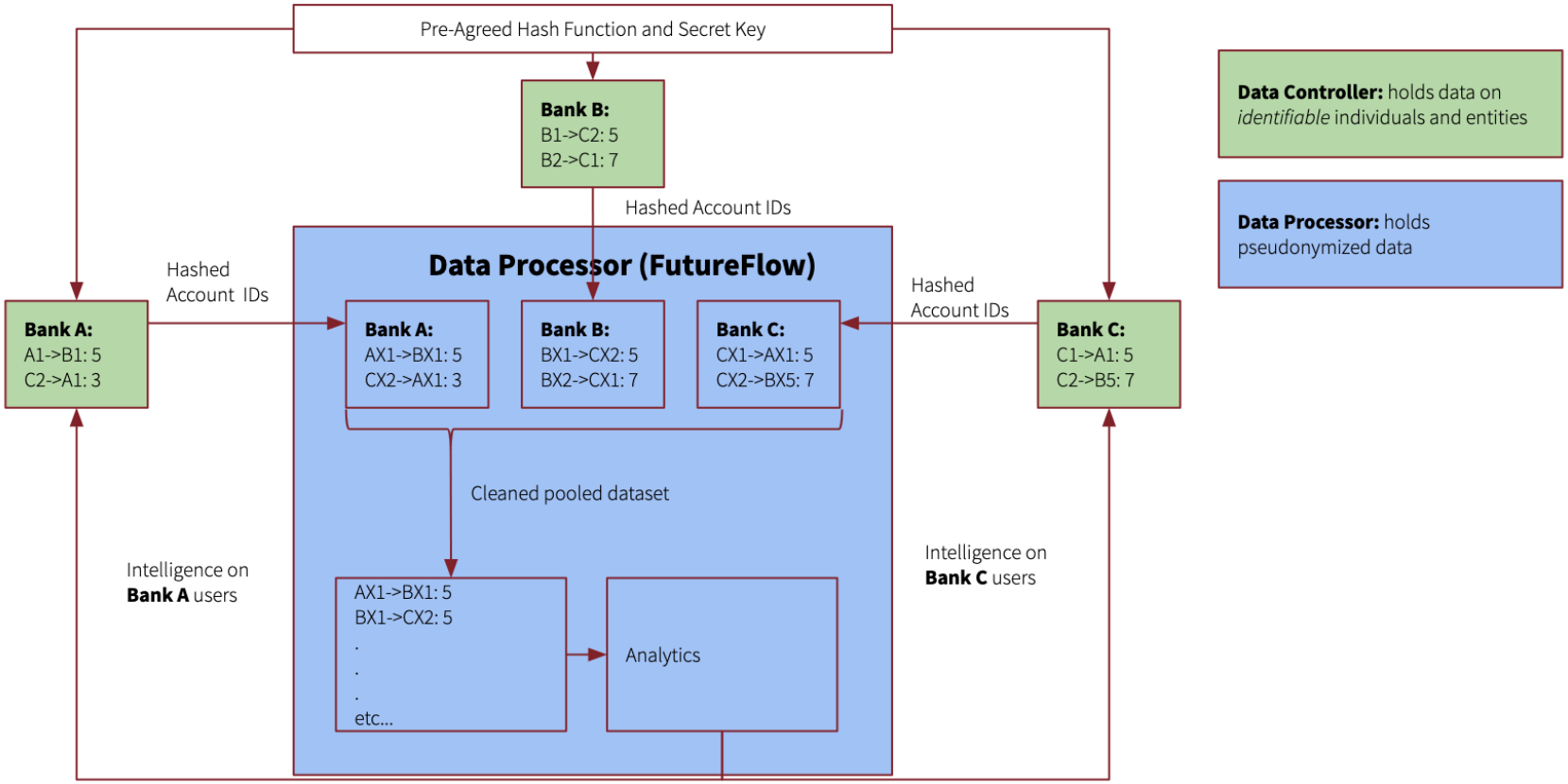
[5] Please Note: the Trusted Third Party will be unaware of the Secret Key.

1. **Deduplication:** for example, a debit transaction in Bank A's dataset, in which a Bank A account sends funds to a Bank B account, may also be submitted as a credit transaction in Bank B's dataset, where the Bank B account receives the funds from the Bank A account. After deduplication, the two transactions from the two separate datasets are represented by one transaction in the pooled dataset where Bank A account sends funds to the Bank B account.

2. **Further Obfuscation (optional):** the account identifiers in the pooled dataset have already been pseudonymized by banks as described above. By pre-agreeing the Hash Function, the Secret Key, and the identifier structure, the participating banks enable the Trusted Third Party to match the same account listed in multiple banks' datasets by its hash without seeing it in plain sight. This enables the Trusted Third Party to replace each account hash in the pooled dataset with a random identifier, since this can be done consistently for each account across the whole pooled dataset. This optional step introduces even further obfuscation of the original account identifier in the final pooled dataset that is sent to FutureFlow for processing, since no backward computational connection remains between the random identifier in the final pooled dataset and the pseudonymised hash that it replaced in the original pooled dataset.

Direct Mode

3.8. In the Direct Mode, each bank or financial institution (ie FutureFlow's clients) submits the transactional data which has undergone pseudonymisation in relation to entity identifiers directly to FutureFlow. In this case, the participating financial institutions agree among themselves on the same synchronisation and pseudonymisation techniques as they would in the Indirect Mode, while FutureFlow performs deduplication and other cleaning operations on the pooled dataset, without relying on the Trusted Third Party. An illustration of how FutureFlow anticipates this transaction will take place is provided below:



3.9. While the Direct Mode reduces the level of complexity in terms of coordination and data exchanges between FutureFlow and the participating financial institutions, it offers a lower degree of data obfuscation compared with the Indirect Mode, since FutureFlow operates on pseudonymised data directly (ie FutureFlow completes the role of the Trusted Third Party as well as its analytics role when operating in Direct Mode).

3.10. Regardless of the Mode chosen by their clients, at the data processing stage, FutureFlow applies its proprietary data transformation and analytics algorithms against the data sets, to map out complex non-linear cross-bank account relationships, extending each participating bank's field of visibility beyond its own boundaries.

3.11. This novel approach to cross-bank financial network mapping and analytics enables multiple banks to tackle financial crime jointly, as opposed to individually using only their own data, by leveraging the combined data of the group. FutureFlow's platform allows their clients to detect where potential malicious actors have deliberately moved their funds among multiple financial institutions to avoid detection, thus empowering financial institutions with a more holistic and in-depth view of the movement of funds across the entire financial system, allowing them to spot at a system-level unusual behaviors and transaction patterns that may constitute financial crime.

3.12. FutureFlow enables two complementary approaches for generating intelligence from its processing:

- **Reactive Approach:** this approach relies on the leading intelligence that can be provided by the financial institutions to generate alerts and insights. For example, some financial institutions may choose to submit transaction flags, seed accounts, or other types of leading intelligence that FutureFlow can use to highlight the relevant parts of the underlying account universe as being strongly associated with the provided accounts.

- **Proactive Approach:** this approach offers automated lead generation, without relying on any leading intelligence supplied by the banks. In this approach, FutureFlow automatically evaluates and ranks the complexity of the generated networks, highlighting those that may deserve particular attention by the Financial Crime analysts of the participating banks.

3.13. The Reactive and Proactive approaches referred to above are complementary and self-reinforcing in enabling the participating financial institutions to conduct transaction monitoring, forensic analytics, and continuous process improvement on the underlying account base. At a minimum, they enable the following use-cases:

- Cross-bank alert generation: enabling blind and automatic cooperation and information sharing regarding problematic

accounts across the participating banks;

- Case triage: enabling each bank to triage alerts and concerns with the benefit of wider cross-bank intelligence; and
- Lead generation: bringing each participating banks' attention to some problem areas in the underlying account base that may have never been discovered before.

3.14. The processing described above is performed at a "pre-suspicion" stage, meaning that financial institutions can submit their transactional data without necessarily having any preconceived knowledge or suspicion of any integrity risks present across the transacting accounts in the submitted dataset. The purpose of the processing is to empower multiple financial institutions to collectively spot, assess, and report criminal transacting patterns by understanding the big picture of the flow of funds.

4. Key data protection considerations

4.1. During FutureFlow's participation in the Sandbox we considered the following key data protection issues:

- Complex data controllership issues surrounding FutureFlow's processing given their different operating models (ie which party should be considered a Controller, Joint Controller or Processor with regard to the processing activity in Indirect Mode and how would this change in Direct Mode?).
- To what extent should the data processed by FutureFlow's platform be considered anonymous and how should FutureFlow manage the risk of reidentification in their platform?
- How should FutureFlow look to comply in a general sense with the UK Data Protection legislation?

4.2. In order to address the above key issues, the ICO and FutureFlow agreed the following four objectives which were documented in FutureFlow's Sandbox Plan:

- **Objective 1:** FutureFlow and the ICO will work to map the ways in which FutureFlow's data could flow given their proposed operating models, and clarify which parties are undertaking key data protection roles in respect of the data utilised by FutureFlow's platform (ie who is a Controller, Joint Controller and/or Processor).
- **Objective 2:** FutureFlow and the ICO will develop a system to effectively quantify and manage the risk of reidentification of personal data in FutureFlow's platform.
- **Objective 3:** FutureFlow and the ICO will develop relevant collateral documentation to help ensure that FutureFlow process personal data in compliance with the UK Data Protection Legislation.
- **Objective 4:** FutureFlow, with guidance from the ICO, should implement sufficient measures to ensure the ongoing security and integrity of the pseudonymous data that they process.

Objective 1 - Complex Data Controllership Issues

4.3. To achieve the first objective the ICO gathered information from FutureFlow about how their platform utilises personal data and the sources from which this data is gathered. The ICO further requested that FutureFlow create visual diagrams which illustrate how the data flowed in order to understand which party ultimately has decision making power over the contributed data. Based on the information provided by FutureFlow it appeared likely that in both the Indirect Mode and Direct Mode of the platform's operation, FutureFlow should be considered a Processor because FutureFlow do not appear to be deciding the means or purpose of the processing activity, nor do they appear to be pursuing their own interests as they are acting in accordance with the instructions of their financial institutions.

4.4. While considering FutureFlow's status as a Processor the ICO also determined that it was likely that FutureFlow's clients (ie banks and other financial institutions) should be considered as individual Controllers for the data they contribute to the pooled dataset, on the basis that they are determining the means and purposes for processing and that they are the party which obtained the data from data subjects in the first place.

4.5. When processing personal data in the Indirect Mode the ICO also considered it likely that any Trusted Third Party working as a middle man between FutureFlow and their clients should also be considered a Processor for their role in the processing activity[6]. This was on the basis that the Trusted Third Party was simply acting under the instruction of financial institutions and was not exercising any real discretion in regards to the processing of the data.

Objective 2 - Anonymous or Pseudonymous Personal Data?

4.6. When considering how best to complete FutureFlow's second objective the ICO had to first consider whether or not the data processed by FutureFlow was anonymous or pseudonymous personal data. This is because Recital 26 of the GDPR makes it clear that the principles of data protection do "…not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."

[6] Please Note: FutureFlow appear to **not** be a Sub Processor for the Trusted Third Party in the Indirect mode of processing as both organisations process personal data on the behalf of the Controller but complete different aspects of the processing.

4.7. Upon examination of the obfuscation techniques used by FutureFlow, their clients and the Trusted Third Parties, it was clear that the methods used would not render the data anonymous as per the above Recital 26 definition and as such the data should be regarded as having undergone pseudonymisation only[7]. The ICO reached this decision because even though the data, once obfuscated and submitted to FutureFlow, could not be identified by FutureFlow there was a chance that a motivated intruder using sufficient measures/techniques could theoretically reidentify the information were they to gain access to it[8].

[7] Pseudonymisation is defined in Article 4,5 of the GDPR as meaning that the data is processed "…in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".

[8] Please Note: FutureFlow's clients (ie the data controllers) will retain access to any secret keys/knowledge of any hashing algorithms used and as such will be able to reidentify the data they receive back from FutureFlow once the processing has taken place. However, controllers will only receive a copy of the data they submitted to FutureFlow or the Trusted Third Party along with any flags on transactions FutureFlow's analytics tool has deemed suspicious.

4.8. In order to manage the risk of reidentification the ICO assisted FutureFlow in creating a suitable model Data Protection Impact Assessment (DPIA)[9] and creating a risk register within this document. The Risk Register effectively documents the risk of reidentification and details the measures FutureFlow have implemented as an organisation (eg security measures, policies and procedures) to mitigate the risk of their clients' data being reidentified.

4.9. Final feedback on FutureFlow's model DPIA has been provided by the ICO and it is now FutureFlow's responsibility to maintain and continue to iterate upon this document as new potential threats and risks emerge.

General Compliance with GDPR

[9] Please Note: Article 35,1 of the GDPR places the responsibility of completing a DPIA on Data Controllers as such it was not strictly necessary for FutureFlow to complete a DPIA, as such the DPIA created by FutureFlow functions as an expanded guide to FutureFlow's platform and is intended to be used as a model document by FutureFlow's clients when they complete their own DPIAs in advance of utilising FutureFlow's platform.

4.10. The last two objectives (3 and 4) of FutureFlow's Sandbox Plan were tackled concurrently and involved the ICO augmenting FutureFlow's data protection expertise by offering ad hoc advice to the organisation and assisting them in drafting suitable collateral documentation to support their ongoing compliance with the UK Data Protection Legislation.

4.11. In particular the ICO advised FutureFlow to create and maintain a Record of Processing Activities (ROPA), Data Protection Policy and Data Protection Impact Assessment (DPIA). The first document created, FutureFlow's ROPA was created by Vadim Sobolevski of FutureFlow, using the data processor version of the ROPA template from the ICO website.

4.12. The ICO worked with FutureFlow between February 2020 and July 2020 to develop, draft and iterate upon their Data Protection Policy and model DPIA. During the course of drafting these documents FutureFlow's Senior Case Officer advised FutureFlow that as a processor of personal data it was not their responsibility to determine an appropriate GDPR Article 6 lawful basis for processing. However, the ICO advised FutureFlow that it was likely that Article 6,1(f) of the GDPR, which states that data processing should be 'necessary for the purposes of the legitimate interests pursued by the controller or by a third party…' was likely to be the most appropriate lawful basis for FutureFlow's clients to rely on in order to share, collate

and commission the analysis of transactional data using the FutureFlow platform[10]. This information was detailed in FutureFlow's model DPIA document.

4.13. During the course of drafting the model DPIA document FutureFlow and the ICO also explored the possibility that FutureFlow's clients may wish to rely on article 6,1(c) of the GDPR ('processing is necessary for compliance with a legal obligation to which the controller is subject') to share, collate and commission the analysis of transactional data from FutureFlow. The ICO advised, in line with existing ICO guidance on the application of article 6,1(c), that, were FutureFlow's clients to rely on this lawful basis for processing, they would have to ensure that the use of FutureFlow's platform was a reasonable and proportionate way of achieving compliance with their specific legal obligation, particularly as it relates to the broad underlying account base, as opposed to just those accounts that may be subject to an existing suspicion. It was deemed unlikely that Article 6,1(c) would be a suitable lawful basis for the clients to process the data, as FutureFlow's platform demonstrates its maximum effectiveness when applied to a broad account base, prior to any firm indication that any accounts have been involved in suspicious activity (ie at the pre-suspicion stage[11]). However, it was deemed that

[10] Please Note - FutureFlow's clients are further advised in the model DPIA document that, if they seek to rely on article 6,1(f) as their lawful basis for processing, that they will be required to complete their own legitimate interests assessment (LIA).
[11] As detailed in the Proactive Approach as outlined in section 3.12 of this document.

FutureFlow's clients may find the article 6,1(c) lawful basis more appropriate in instances where the suspicious cross-bank relationships and phenomena uncovered by the FutureFlow processing are considered to constitute sufficient evidence of suspicion, so as to require a closer joint investigation by the involved banks[12]. To be clear, any further processing in such cases would not involve FutureFlow as data processor.[13]

4.14. As part of FutureFlow's participation in the Sandbox the ICO provided some advice to the organisation on the security section of their Data Protection Policy. This feedback was provided to FutureFlow on29 May 2020, and required FutureFlow to take certain action accordingly. In June 2020 the ICO and FutureFlow continued to work together to improve the substance of FutureFlow's model DPIA, focusing in particular on the wording and format of their risk assessment section. At the time that this exit report was written FutureFlow, had received comments on several iterations of their DPIA and Data Protection Policy. It is now FutureFlow's responsibility to maintain and continue to iterate upon these documents to ensure that they remain an accurate and up to date reflection of the relevant data protection risks and mitigations.

---

[12] As detailed in the Reactive Approach as outlined in section 3.12 of this document.
[13] Please Note – although the ICO has provided advice to FutureFlow regarding which lawful basis their clients may wish to rely on it is ultimately up to FutureFlow's clients (ie the Data Controllers) to make this decision themselves.

4.15. The ICO also worked with FutureFlow to create some suitable template wording, for inclusion in their clients privacy notices, regarding how FutureFlow will process personal data.

5. Ending statement

5.1. FutureFlow's participation in the ICO's Regulatory Sandbox has given the ICO the opportunity to gain a valuable insight into the financial sector and how banks and other financial institutions might choose to leverage and share the data they already collect to detect and tackle instances of financial crime. The ICO's work with FutureFlow will help to influence our views and any future work on how large organisations can anonymise, pseudonymise and share data for the purposes of tackling financial crime in a compliant and secure manner while maintaining individuals' rights to privacy.

5.2. It is clear to us from our work with FutureFlow that they have a real commitment to making use of innovative technology in a compliant way to improve the way banks and other financial institutions collaborate when addressing the problem of financial crime. These include ensuring that:

- compliant contracts, as required by Article 28 of the GDPR, or data sharing agreements are in place between all parties involved in the processing;
- DPIAs are completed in advance of the processing activity taking place which take into account the inherent risk of reidentification and detail the methods which will be used by the parties involved to minimise this risk; and
- reports which indicate financial crime may have taken place are properly investigated, where corresponding financial legislation allows, before suspects experience any changes to their level of service (ie suspects do not have their banking services removed purely on the basis of a report by a platform like FutureFlow's).

5.3. Furthermore, based on the information we have seen in the Sandbox and solely in respect of the FutureFlow platform as considered in the Sandbox, it appears likely that the financial transactional data processed by FutureFlow's platform is processed securely and not in a way which breaches UK data protection legislation. Moving forwards FutureFlow should ensure that they continue to follow the steers provided to them in the Sandbox, as well as relevant ICO guidance.

5.4. At the time of writing this report the Covid-19 public health emergency was unfolding across the world with unprecedented impact on the financial sector. We are immensely grateful to FutureFlow for their engagement in the Sandbox and finalising this report in difficult circumstances.